Mini Project

Problem Statement

BigMart is a large retail chain with stores in multiple cities. The company has collected sales data for 1559 products across 10 stores in different cities. The data includes information about the product attributes (e.g., product type, brand, size, weight) and store attributes (e.g., location, size, number of employees).

The goal of this project is to build a predictive model that can be used to predict the sales of each product at a particular store. The model will be used by BigMart to improve its inventory management and marketing strategies.

Specific Objectives

The specific objectives of this project are to:

* Clean and prepare the data for analysis
* Identify the most important product and store attributes that affect sales
* Select an appropriate machine learning algorithm to build the predictive model
* Evaluate the performance of the model
* Deploy the model into production

Hypothesis Generation

* Hypothesis 1: The sales of a product at a particular store are affected by the product type. For example, products that are frequently purchased, such as food and beverages, may have higher sales than products that are less frequently purchased, such as electronics.
* Hypothesis 2: The sales of a product at a particular store are affected by the store location. For example, stores that are located in high-traffic areas may have higher sales than stores that are located in less-trafficked areas.
* Hypothesis 3: The sales of a product at a particular store are affected by the store size. For example, larger stores may have higher sales than smaller stores because they have more shelf space and can stock a wider variety of products.
* Hypothesis 4: The sales of a product at a particular store are affected by the store's marketing efforts. For example, stores that run more advertising campaigns may have higher sales than stores that do not.

These are just a few examples of hypotheses that could be predicted for the BigMart sales prediction problem statement. The specific hypotheses that are generated will depend on the specific data that is available. However, these hypotheses provide a starting point for us while working on this project.